DIGITAL AUDIO FORENSIC

Acquisition, Analysis, Authenticity, Audio Enhancement, Audio Interpretation

Manas Kumar Mishra
Dept. of Electronics and Communication Engineering
Indian Institute of Information Technology, Design and Manufacturing
Kancheepuram, India
esd18i011@iiitdm.ac.in

***Abstract*—This document is a lab report for ADSP lab. In that, I am discussing the Digital Audio Forensic process. I have written the codes in matlab completely. In the discussion, I am representing various algorithms and methods for Audio authenticity test, Audio enhancement and Audio Interpretation. In all sections, I have provided a brief subsections for different methods with results in matlab. Along with I have represented block diagrams or brief steps for every algorithm. I am following the order followed by the experts to analyze the audio. At the end, I have tested with seven Indian audio clips recorded by ordinary people in various Indian scenarios.**

***Index Terms*—Audio Forensic, Audio authenticity, MFCC, Audio Enhancement, Audio Interpretation, Compressor, Multilateration**

I. Introduction

A. Acoustic

Sound is the disturbance that moves away from the source such that direction of oscillation and propagation of the disturbance are same. Basically, it is a type of longitudinal wave and study of sound is known as *Acoustic*. Acoustic, science of sound contains Production, Control, Transmission, Reception and Effects related to the all types of sound. That makes acoustic point of view, as a very important aspect for any sound related arguments.

B. Audio Forensic

Audio forensic is a special type of study of sound based on acoustic and digital signal processing principles such that court can use that study as an evidence and based on that it can sanction the result. In reality, Audio forensic study depends upon the type of case, court direction for that case and demand of lawyer from the audio analysis. Hence, we can not generalize step by step procedures for audio forensic. But in a broader sense, we can divide audio forensic study into three buckets, *Authenticity, Audio Enhancement, Audio Interpretation* [1].

Thanks for all classmate, lab TAs and Asutosh kar sir for their support and opportunity.

C. Why? Digital Audio forensic

Because, 2.5 Quintilian bytes of data are created by human each day for communication and storage purposes. Due to internet, Social media platforms and sophisticated smart devices, this amount will increase more, in upcoming future [3]. In this huge amount of data audio plays an important role from normal TV ads to normal phone calls, social media video post to fancy cinema songs.
But now we can see the increasing issues and crimes in the digital data domain, For that we need a separate type of analysis and rules for this domain. Even in current situation, court need to check the audio evidences for a crime or potential crime. In that case, experts have to dig down into the digital equivalent of real event to extract or proof something. That makes Digital Audio forensic important for future and present as well.

II. Evidence Handling And Initial Assessment

Before, starting the any type of forensic analysis on audio recording, forensic experts have to note the basic audio characteristic of original audio recording, initially assuming that audio recording is untempered and true audio. For that they use basic digital signal processing techniques and critical listening for that audio [1].
That helps them to make final report for court proceedings.

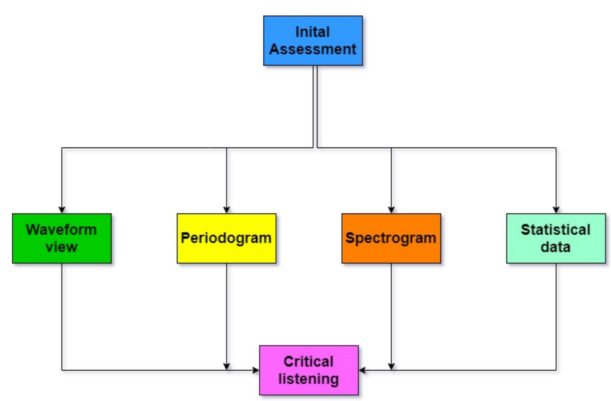

Fig. 1. Initial Assessment

A. Waveform view

It just a amplitude and time representation of the recorded audio.

$$f[n] = y \tag{1}$$

Where, n is sample value, y is amplitude corresponding to that sample and f is a function that maps the sample value to amplitude value. From this, experts can extract time domain features of audio recording. The temporal features (time domain features), which are simple to extract and have easy physical interpretation, like: the energy of signal, zero crossing rate, sampling rate, maximum amplitude, minimum amplitude, etc. Experts notes all this things for future reference.

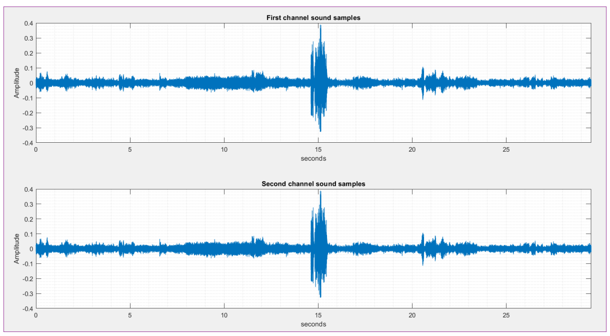

Fig. 2. Waveform view of audio signal (Test1_my_a.wav)

In my case sampling rate is 48000 Hz.

B. Periodogram and power spectrum

Periodogram is a method for finding the power spectral density (PSD). In reality experts use the another specific software than Matlab for see the PSD, in case of matlab, I was getting very unclear PSD by periodogram method, hence I used another inbuilt method for PSD estimation (pspectrum).
First, With Periodogram, I have designed the code such that, Matlab ask a quest to choose the window like this.

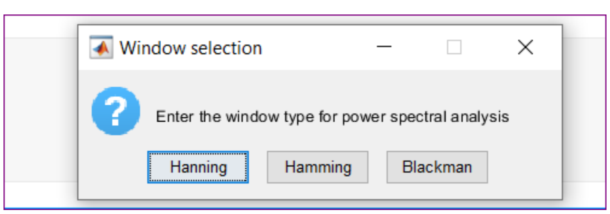

Fig. 3. Quest for window method

In another function I was using pspectrum for PSD calculation output is like this

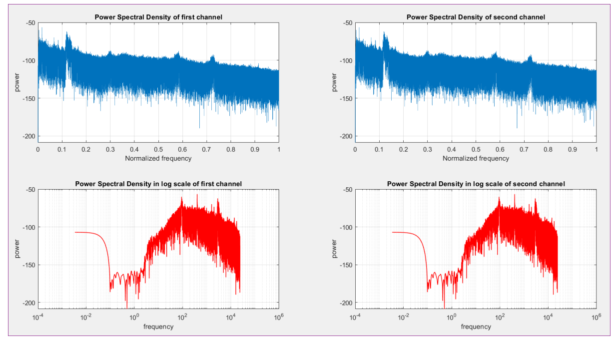

Fig. 4. PSD by periodogram

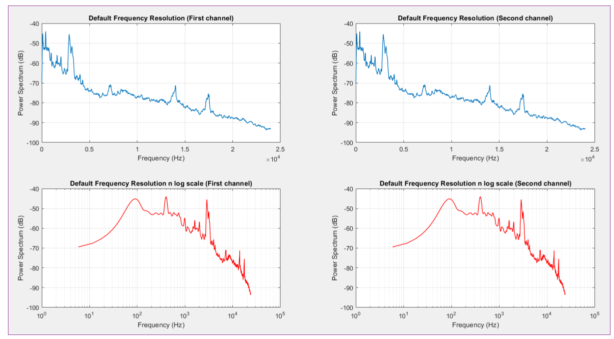

Fig. 5. PSD by pspectrum

C. Spectrogram

In spectrogram, we take time and frequency information together for representing the audio. First, we divide the audio samples into small segments, by applying windowing and fft on each segments with small overlap region, we obtain the *STFT* Short Time Fourier Transform. At the end we apply log transform on *STFT*, that give us log scale *STFT*. We represent that into 2D plane having time and frequency as x-axis and y-axis respectively with color showing energy information and that x and y position.

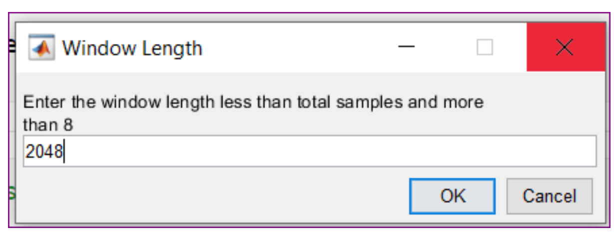

Fig. 6. Window length input

It looks like an image, hence it is often known as image representation of audio signal. But the problem is in the choice of window length for proper frequency and time resolution. For long window size, there is good frequency resolution, but for short window size, there is good time resolution. Hence, we need to choose different window size

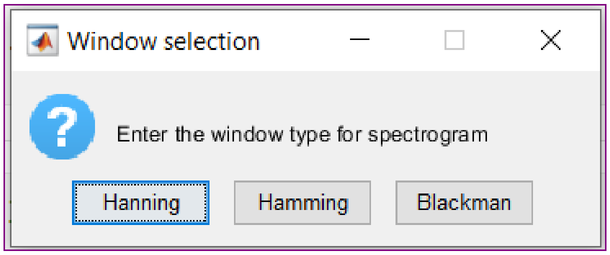

Fig. 7. Asking type of window

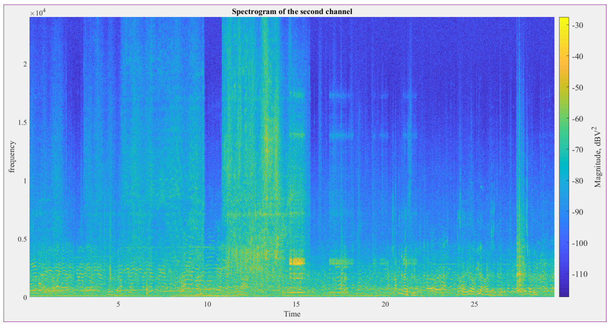

Fig. 8. Spectrogram (window Length = 2048, Window type = Hamming window)

for different applications.

D. Statistical data

Here, we consider the maximum and minimum values, mean, variance, and standard deviation of the signal, and power of the signal in dB.

```
Max value in first Channel = 0.38922
Max value in second Channel = 0.19986
Minimum value in first Channel = -0.32739
Minimum value in second Channel = -0.18192
Mean value in first channel = -3.2807e-06
Mean value in second channel= -3.147e-06
standard deviation of first channel= 0.017072
standard deviation of second channel= 0.014982
Variance of first channel= 0.00029145
Variance of second channel= 0.00022447
Power in first channel in dB= -35.3544
Power in second channel in dB= -36.4884
Sampling frequency in Hz =48000
```

Fig. 9. Audio statistical data

At the end experts have to listen the original audio multiple time such that they can answer the questions asked by court related to that audio recording or evidence.

III. Authenticity Assessment

The main purpose of this to educate the court about doubts on originality of audio, define the circumstances in that recording was made, is there any sign of dispute about the recording and it's authenticity.In nutshell,

1. Is recording complete?
2. Is recording unaltered?
3. Is recording consistent?

A. First order authenticity test

In this test, experts check the container of audio data and various waveform of the audio, but that is not a solid testing techniques, because anybody with the little bit expertise can overcome this test.
In the container test, they check the file format, file header, MAC (Modification time, Access time, and Creation time) time stamps, hex data of the file. Basically, all possible parameters linked with audio file that can show the clue of any editing inside the file. But as we know, now a days, we can easily create a new file after editing the audio samples and then delete or permanently replace the original file by tampered file.

In waveform authenticity analysis, we check the discontinuities in spectrogram that shows the potential deletion or addition in the energy level of the audio.

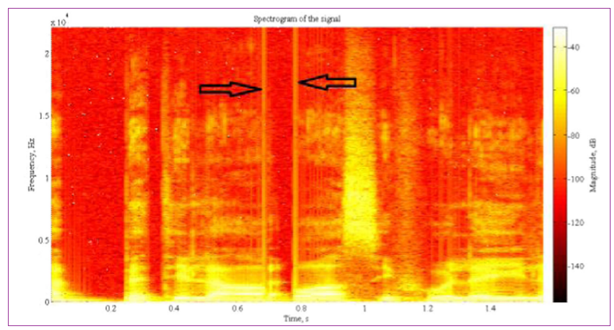

Fig. 10. Abrupt change in the spectrogram potential sign of editing [2]

But this is not a solid scientific proof and any-body can overcome this by careful editing in the audio.

B. Higher order authenticity test

In higher order authenticity test they use, data driven approach for example Authenticity using environment signature and microphone signature [6]. In these approaches, they use big recorded data-sets and trained algorithms for test the new data or feature of audio and classify it appropriately. That allows us to point out the different audio environments or different microphone regions in recorded audio, a scientific proof of editing. Full process is beyond the digital signal processing domain.

But, we can talk about the feature extraction from the audio. I have made a function to extract the *MFCC mel-frequency cepstral coffecients* using inbuilt functions in matlab. Similarly, we can extract the cepstral features and mel-scale spectrogram.

1) Mel-spectrum: We, humans perceive the audio in logscale. The audio perceived by human ear from 400Hz to 600Hz is much different than 1400 Hz to 1600Hz even at same pressure level. Hence, we have mel-scale frequency

bands where mel is derived from melody.
Mel-spectrogram is the normal spectrogram with frequencies mapped to the mel-frequencies. Here, we convert the actual frequency of STFT into mel-scale using given formula

$$m = 2595.\log_{10}\left(1 + \frac{f_{Hz}}{const}\right) \tag{2}$$

Where const is 500 or 700 depend upon the designer. I have used matlab to plot the mel-spectrum with histogram. Researcher designed the model to learn the features from this representation.

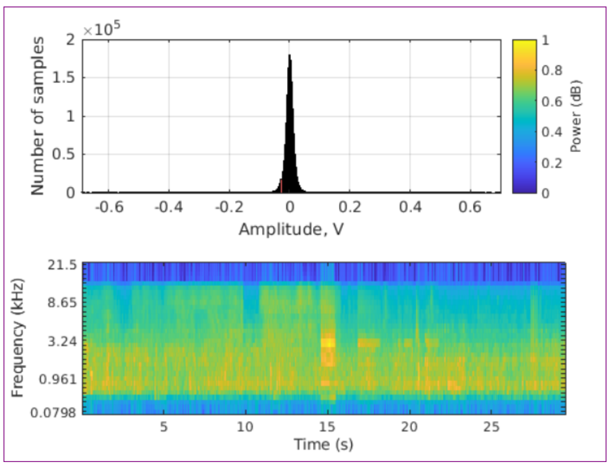

Fig. 11. Mel-spectrogram and histogram

2) MFCC: It is a very special type of feature that is important for many type of speech processing application. Block diagram is given below.

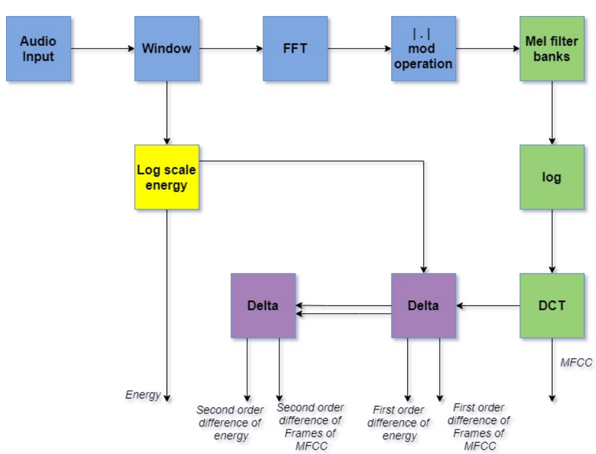

Fig. 12. Block diagram of MFCC

12 is the dimension of MFCC and it's first and second order difference in matlab. But for energy it is just a scalar quantity that implies maximum overall feature dimension is 39. From these feature vectors we can apply the many classifier algorithm to classify and authenticate or recognize the pattern in the audio evidence [6], [7].
There are many other features and descriptors present like MPEG7 descriptors, TDSM based features.

IV. Audio Signal Enhancement

To improve the intelligibility of the conversation from a poor quality surveillance recording obtained with a hidden microphone. Improving signal quality and intelligibility is subjective to the listener, type of recording device and type of noise.
There are many solutions exist, but main problem which technique should we use for a particular audio evidence. Here, I am showing band-pass filtering, wiener filter and compressor for enhancing the audio. In reality, we have many like AGC (Automatic Gain Control), spectral subtraction.

A. A basic filter for vocal frequencies

As we can see in figure 5, there are many frequency components are involved in any audio evidence, because it has not been recorded in lab or by any expert. That audio evidence is just a piece of digital equivalent of real event and recorded through day to day microphones or recording devices. Almost all devices are designed for catching frequencies from 20 Hz to 20 KHz (Audible frequency range). But human vocal cord generates sound from 200 Hz to 4 KHz. Hence, we can apply the band-pass filter to reduce the extra frequencies for audio enhancement (Speech intelligibility).
In my case, order of the filter is 256 and it's response is given like this. Output is given in fig 14

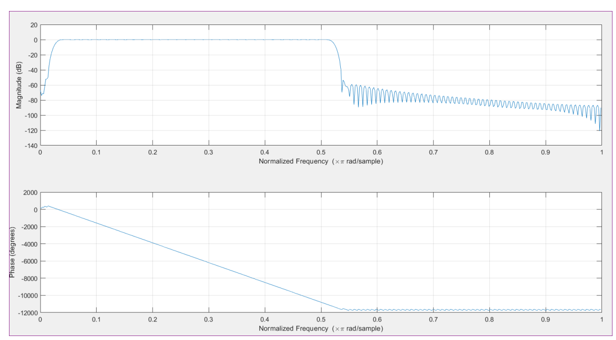

Fig. 13. Filter response

B. Wiener filter (Outside code help)

After the Vocal filtering, I am applying wiener filter to reduce the noise drastically [5]. But, for that I am using small parts of outside code by *Esfandiar Zavarehei* [4](not written by me completely).
But it is working amazingly for all audio clip to increase speech intelligibility. The filter domain output is given in fig

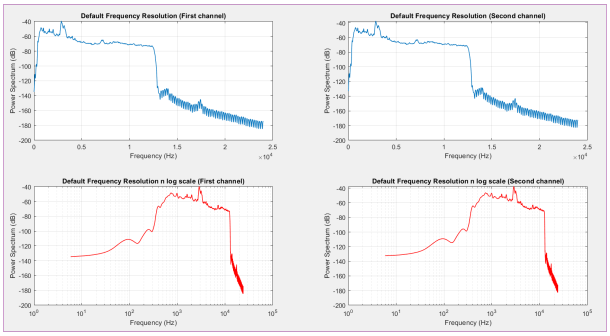

Fig. 14. PSD after filtering

15 and block diagram is given in fig 16

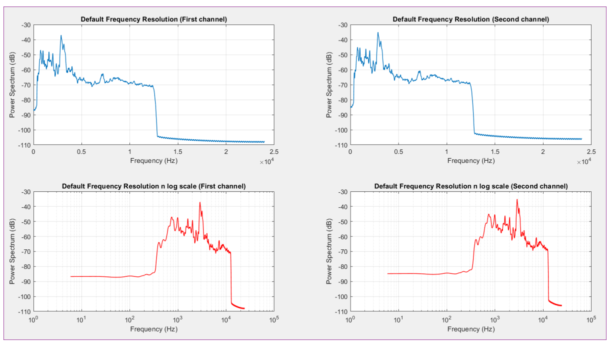

Fig. 15. PSD after Wiener filtering

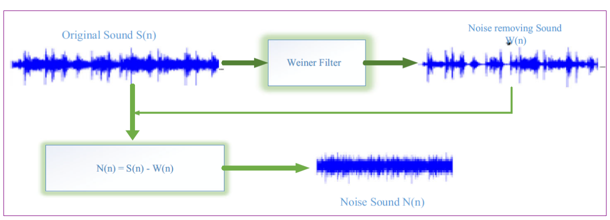

Fig. 16. Wiener filter operation

C. *Compressor or Expander*

After the removing the noise, I am applying compressing or expanding techniques for more understanding of sound. In the compressor, it will compress the sound of high power and allow low power sound as it is. The characteristic of compressor is given in fig 17.

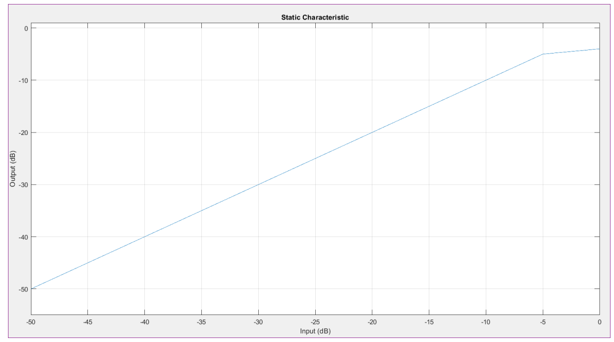

Fig. 17. Compressor Characteristic

V. Audio Interpretation

After finding authenticity and enhancing the audio, experts has to answer the court questions about the recorded event. For that they have to apply acoustic or physics principles apart from DSP to interpret the audio and proof or disproof arguments of prosecution and defender. That methods used by experts for this purpose should be scientifically proven because here stacks are very high. Here, I can not design matlab program for all possible audio, but I am giving a brief overview of some techniques used by experts and potentially can be design by matlab.

A. *Multilateration*

Let's suppose we have at least two audio recordings of same event but from different microphones at different locations. First, we have to synchronize them. Then we need to estimate the location of the sound source. This whole process is known as *Multilateration*.
It model is simple, we know the locations of the mic and speed of sound at that temperature as a prior knowledge. Then we can model as given below with L + c dt and L - c dt.

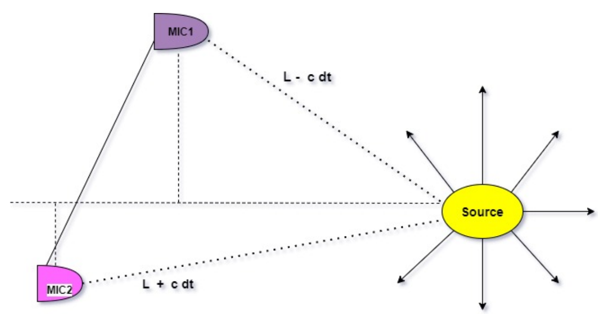

Fig. 18. Multilateration Model

Main idea is that difference between both path is constant that is nothing but locus of hyperbola

$$|path|_1 - |path|_2 = 2cdt \quad (3)$$

Due to this we can estimate just the possibility of the actual location about the both mic positions. That is enough for proof the prosecutor or defendant arguments. But wait a second, the main question is how to find time difference? And what is that time difference?
Here, time difference is the time gap between the sound event in both recordings after synchronization. For example, a gun short is recorded by one microphone at 20 sec after started

it's recording, but same sound is recorded in second mic at 22 sec at same started time (synchronization). Here time gap is 2 sec.

For finding this time gap we use cross correlation operation between both sound samples. The gap between the peak of cross correlation output and y axis indicates this time gap with sampling rate. Here we use matlab to determine this gap.

Somehow, if we get third or more recordings with their mic locations then we can exactly point out the actual position of source by finding the intersection of two or more locus or loci.

B. Doppler effect

Most of the cases, experts need to define the speed of the source object or speed of recording device. In these type of cases, Doppler formula is most reliable scientific way. But the issue is speed of sound is varying with respect to temperature and how to find actual frequency of sound source because, due to Doppler's effect we know that recorded frequency would be different than actual frequency. For these sort of questions in audio forensics, we need to see the manufacturing sides of the sound source and many other techniques.

C. Likelihood test

It is a nice scientific way to model the prosecution and defendant side arguments. And represent the result of test in scientific ways, like in which side result is pointing truth.

$$Ratio\ of\ probabilities = \frac{(Prosecution\ Hypo|test\ result)}{(Defendant\ Hypo|test\ result)} \tag{4}$$

From this, we can model the reliability of the test also.

Conclusion

Digital Audio forensic is a very big field in itself, I have touch just a small part related to ADSP topics. Authenticity and Enhancement are very subjective in nature and Interpretation is full depend upon the query asked by court. But for this lab case, I have tested my project with Seven real audio recordings.

References

[1] [Modern Acoustics and Signal Processing] Robert C. Maher - Principles of Forensic Audio Analysis (2018, Springer International Publishing).

[2] Maher RC (2010) Overview of audio forensics. In: Intelligent multimedia analysis for security applications. Springer, vol. 282, pp. 127–144.

[3] https://seedscientific.com/how-much-data-is-created-every-day/

[4] https://ww2.mathworks.cn/matlabcentral/fileexchange/7673-wiener-filter?$s_tid = FX_rc1_behav$.

[5] Moon C-B, Kim H, Kim BM (2014) Audio recorder identification using reduced noise features. In: Ubiquitous information technologies and applications, Springer, pp 35–42

[6] Zakariah, M., Khan, M.K. & Malik, H. Digital multimedia audio forensics: past, present and future. Multimed Tools Apply 77, 1009-1040 (2018).

[7] Rabiner, Lawrence R., and Ronald W. Schafer. Theory and Applications of Digital Speech Processing. Upper Saddle River, NJ: Pearson, 2010.